2021

Application of Machine Learning Classification Methods in Fault Detection and Diagnosis of Rooftop Units

Amir Ebrahimifakhar
University of Nebraska-Lincoln, United States of America, aebrahimifakhar@unomaha.edu

David Yuill

Adel Kabirikopaei

Application of Machine Learning Classification Methods in Fault Detection and Diagnosis of Rooftop Units

Amir EBRAHIMIFAKHAR[1]*, David YUILL[1], Adel KABIRIKOPAEI[1]

[1]University of Nebraska - Lincoln, Architectural Engineering,
Omaha, NE, USA
amir.ebrahimifakhar@huskers.unl.edu, dyuill@unl.edu, adel@huskers.unl.edu

* Corresponding Author

ABSTRACT

In this paper, a data-driven strategy for fault detection and diagnosis in rooftop air conditioning units, based on machine learning classification methods, is proposed. The strategy formulates the fault detection and diagnosis task as a multi-class classification problem. The focus of this study is on detecting and diagnosing the following common rooftop unit faults: refrigerant undercharge, refrigerant overcharge, compressor valve leakage, liquid-line restriction, condenser fouling, evaporator fouling, and non-condensable gas in the refrigerant. Three classification methods, K-nearest neighbors, logistic regression, and random forests were applied to our dataset, and their performance was compared. Ten-fold cross-validation was used to select tuning parameters for different classification methods. Machine learning requires a larger set of training data than could feasibly be generated with experiments, so a library of high-fidelity simulation data was used to train and test the classification methods' performance. This library contains physical parameters – such as temperatures, pressures, and power consumption – for a three-ton rooftop unit operating with each of the fault types individually, at several fault intensities, over a range of driving conditions. The logistic regression method, with an overall accuracy rate of 93.6%, had the best performance of the methods that were tested. Since class-specific performance is also important in fault detection and diagnosis process, the performance of different classification methods for individual faults was also determined, using true positive rate and false positive rate statistical measures. The results demonstrate the potential of data-driven strategies to detect and diagnose common rooftop unit faults.

1. INTRODUCTION

Air conditioning equipment uses a significant portion of the energy in the US. In commercial buildings, particularly small and medium sized buildings, packaged rooftop unit (RTU) systems are the most common form of air-conditioning system (EIA, 2012). These systems are simpler for building design and construction professionals to apply than many alternate systems (e.g. chiller, boiler, piping, pumps, air handling units, variable air volume distribution system). In large settings, such as big box stores, a large load is often met by specifying multiple RTU. For example, a 300-ton load could be met with 15 RTU, each of 20 tons capacity. Often, the maintenance effort for a building is therefore spread out among many systems, each of which is expendable without seriously compromising indoor comfort. As a result, it is believed that RTU may receive less maintenance attention than other types of systems. This means that degradation faults can go unnoticed for long periods of time, resulting in energy waste (Shoukas *et al.*, 2020). The faults may also add further operating penalties by reducing the lifespan of the equipment (Li and Braun, 2007d; Yuill and Braun, 2017).

Fault Detection and Diagnosis (FDD) tools can help to identify faults early, so that they can be addressed before wasting significant energy or equipment life (Braun, 1999; Li and Braun, 2007b; Yuill and Braun, 2016).

There are several studies that propose FDD methods for RTU systems. Rossi and Braun (1997) developed a statistical, rule-based FDD method for vapor compression air conditioners with single-stage compressors, fixed-speed fans, and fixed orifice expansion valves. Chen and Braun (2001) modified the original FDD approach proposed by Rossi and Braun (1997) for RTU systems having a thermostatic expansion valve (TXV) as the expansion device. Their algorithm was designed to detect and diagnose seven faults in RTUs. Armstrong *et al.* (2006) developed an electrical-signal-based FDD technique for RTU systems. Li and Braun (2007a, 2007c, 2009a, 2009b) developed a decoupling-based FDD method that can handle multiple simultaneous faults for packaged air conditioners. Ebrahimifakhar *et al.* (2020) developed an FDD approach for RTUs using machine learning classification methods.

Most FDD approaches in the literature that were developed for packaged air conditioners are rule based, meaning that a set of expert rules that relate to physical traits of the system are applied to measurement data. Rules can be applied to operating parameters, such as suction superheat, or to observable behaviors, such as energy consumption. Many rule-based FDD tools are effective, but there are challenges. Yuill and Braun (2012, 2013) documented a method for testing the performance of FDD tools, and found that performance was lacking for many of the tools that they tested. A common challenge for the FDD methods was distinguishing between faulted operation and unfaulted operation in an unusual operating condition. Typical rule-based methods apply a threshold that must be exceeded, and selecting this threshold appropriately is difficult. Another challenge was that tools misdiagnosed fault types. Similar to false positives, one fault type combined with an unusual operating condition could cause rules to be broken in a way that was unanticipated by the rule developer.

Data-driven methods have recently been developed for other types of systems such as chillers (Han *et al.*, 2011; Zhao *et al.*, 2013) and air handling units (Yan *et al.*, 2016; Yun *et al.*, 2021). However, very few data-driven methods have been developed for RTU (Ebrahimifakhar *et al.*, 2020). There may be significant potential for the use of recently developed approaches for machine learning, to develop data-driven FDD algorithms for RTU systems. These methods require large sets of data for training. The hope is that they eventually could augment or even outperform existing FDD methods in ease of development, ease of deployment, or accuracy. However, currently the availability of sufficient training data poses a significant barrier to development. A single set of publicly available data from fault tests of a chiller conducted by Comstock *et al.* (2001) has been used by many researchers to develop data-driven methods, but it is not clear whether data-driven methods could be effective on air-cooled unitary equipment, such as RTU and split systems.

The purpose of the current research, therefore, is to conduct a preliminary investigation to better understand the potential of machine-learning for RTU as an effective classification method, to differentiate faulted from unfaulted conditions, and to differentiate fault types. It is the first step in understanding the potential of these methods. If machine-learning methods are sufficiently promising, for a single air-conditioner, additional research questions to be answered include: How does their performance vary from system to system? Which classifiers work best? Is there potential to train classifiers on one system and apply them to other systems? What happens if there are multiple simultaneous faults? How does diagnostic performance trade off against reduction of the number or type of measurement inputs? How can we generate sufficient training data in a cost-effective manner? What are the minimum requirements for these training data? We envision a potential future scenario in which RTU manufacturers generate training data from component-based models, and use these as a training set for onboard data-driven diagnostics.

This paper presents a data-driven FDD strategy for RTU systems using statistical machine learning classification methods. Three classification methods are applied to our dataset in order to detect and diagnose the faults, and the accuracy of the methods is analyzed. The high accuracy of the classification models in detecting and diagnosing the faults shows that the proposed approach is promising.

2. METHODOLOGY

The focus of this research is on detecting and diagnosing of the RTU faults that degrade the system performance but permit continued operation of the system. The following fault types are considered: (1) refrigerant undercharge (UC), (2) refrigerant overcharge (OC), (3) compressor valve leakage (VL), (4) liquid-line restriction (LL), (5) condenser fouling (CA), (6) evaporator fouling (EA), and (7) the presence of non-condensable gas in the refrigerant (NC). The data-driven approach, which is presented in this study, simultaneously detects and diagnoses the faults in a single step, using machine learning classification methods. The classification problem, in other words, includes the potential for classification into the unfaulted class, or into one of the fault classes. Over the past few decades, many classification algorithms have been developed, and are widely used in FDD applications for a large variety of engineering applications (Bishop, 2006; Fernández-Delgado *et al.*, 2014; James *et al.*, 2013). K-nearest neighbors (KNN), logistic regression (LR), and random forests (RF) were selected as the three machine learning algorithms to be applied in this study.

Experimental data for RTU systems is rare and costly to generate. The data need to be sufficiently reliable and accurate that they need to come from careful laboratory testing. The dynamics of air-conditioning systems can be very complex, so the problem is simplified by limiting the RTU fault classification to systems under steady operation. Nevertheless, conducting enough steady-state tests to train a machine learning algorithm could take years. Therefore, this study relies on a data library that was generated from a simulation by Cheung and Braun (2013a, 2013b). In this dataset inverse modeling has been used in a component-based grey box model to generate system performance properties under faulted and normal (no fault) conditions for an RTU. This model was validated for use in evaluating the performance of FDD tools by Yuill *et al*. (2014).

A three-ton RTU is used as the subject for testing the proposed data-driven FDD strategy. The specifications of the RTU are shown in Table 1.

Table 1: Specifications of the RTU

Nominal Capacity [kW]	Refrigerant	Expansion Device	Condenser Type	Compressor Type	Operating Mode
10.6	R410A	Fixed Orifice (FXO)	Fin-tube	Scroll	Cooling

The thermodynamic state of the RTU system under both normal and faulted conditions is characterized by the following variables (features): (1) return air dry bulb temperature (T_{RA}), (2) return air wet bulb temperature (WB_{RA}), (3) supply air dry bulb temperature (T_{SA}), (4) supply air wet bulb temperature (WB_{SA}), (5) ambient air dry bulb temperature (T_{amb}), (6) liquid-line pressure (P_{LL}), (7) liquid-line temperature (T_{LL}), (8) suction pressure (P_{suc}), (9) suction temperature (T_{suc}), (10) compressor discharge pressure (P_{dischg}), (11) compressor discharge temperature (T_{dischg}), (12) condenser exiting air temperature ($T_{air,ce}$), (13) refrigerant saturation temperature in the evaporator ($T_{sat,e}$), (14) refrigerant saturation temperature in the condenser ($T_{sat,c}$), and (15) compressor power ($Power_{comp}$). Return air (features 1 and 2) refers to the air entering the evaporator. Statistical descriptors of the input variables in the dataset are shown in Table 2.

Table 2: The statistics of the input variables

Input Variable	Unit	Mean	Standard Deviation
T_{RA}	°C	25.0	2.9
WB_{RA}	°C	17.6	4.3
T_{SA}	°C	14.9	3.7
WB_{SA}	°C	12.2	4.7
T_{amb}	°C	32.4	9.0
P_{LL}	kPa	2748.2	619.1
T_{LL}	°C	38.4	10.4
P_{suc}	kPa	1031.0	129.2
T_{suc}	°C	13.2	6.1
P_{dischg}	kPa	2842.7	613.2
T_{dischg}	°C	71.6	16.6
$T_{air,ce}$	°C	42.8	9.4
$T_{sat,e}$	°C	8.1	4.6
$T_{sat,c}$	°C	44.5	9.6
$Power_{comp}$	W	2527.1	639.1

The dataset contains the fifteen input variables for each of the 2851 unique observations (samples). The output variable takes on one of eight possible categorical values: (1) UC, (2) OC, (3) VL, (4) LL, (5) CA, (6) EA, (7) NC, and (8) NF (no fault). Due to the different units and ranges of the input variables, during data conditioning the data were standardized so that each variable has a mean value of zero and a standard deviation of one. The severity of each fault is characterized with fault intensity (FI), as defined by Yuill and Braun (2013), and summarized in Table 3. FI is an indicator of the severity of the fault with respect to directly measurable quantities. The grid of fault intensities within the training dataset is also shown in this table.

The 2851 observations were divided into a training set and a test set. The training dataset includes 2/3 of the points (1901 observations) and test dataset consists of the 950 remaining observations. Table 4 shows the name of the dataset, number of observations, and quantity within each class in the dataset. K-nearest neighbors (KNN), logistic regression (LR), and random forests (RF) are used as the three classic machine learning algorithms in this study. All three statistical models were implemented using the popular statistical software R (2019). Each model was trained using the training dataset and its performance was evaluated on the test dataset. The higher the accuracy on the test dataset, the better the performance of the classifier. Each method may have several tuning parameters. Tuning parameters are usually chosen using the k-fold cross-validation (CV) technique (Bishop, 2006; James *et al.*, 2013). In this study, a 10-fold CV technique was used to choose the tuning parameter values for the KNN and LR methods. Out-Of-Bag (OOB) accuracy was used for the RF method to choose appropriate values for the tuning parameters (James *et al.*, 2013).

Table 3: Fault intensity definitions and values

Fault type	Fault intensity definition	Fault intensity values (%)
UC or OC	$FI_{charge} = \frac{m_{actual} - m_{nominal}}{m_{nominal}}$	70, 80, 90, 110, 120, 130
VL	$FI_{VL} = \frac{\dot{m}_{faulted} - \dot{m}_{unfaulted}}{\dot{m}_{unfaulted}}$	10, 20, 35, 50
LL	$FI_{LL} = \frac{\Delta P_{LL,faulted} - \Delta P_{LL,unfaulted}}{\Delta P_{LL,unfaulted}}$	50, 100, 300, 600, 1200, 2000, 3500
CA	$FI_{CA} = \frac{\dot{V}_{actual} - \dot{V}_{nominal}}{\dot{V}_{nominal}}$	90, 77, 63, 50, 40
EA	$FI_{EA} = \frac{\dot{V}_{actual} - \dot{V}_{nominal}}{\dot{V}_{nominal}}$	90, 75, 60, 45
NC	$FI_{NC} = \frac{m_{N_2,faulted}}{m_{nominal}}$	10, 30, 55, 80, 100

Table 4: Main characteristics of the datasets

Dataset	Observations	UC	OC	VL	LL	CA	EA	NC	NF
Training Data	1901	252	353	367	233	263	239	163	31
Test Data	950	146	159	166	103	143	127	89	17

Finally, to evaluate and better understand the performance of the models in detecting and diagnosing the faults, confusion matrices were used. The number of correct predictions (the diagonal elements of the confusion matrix) divided by the total number of observations is used as the Overall Accuracy Rate (OAR). The True Positive Rate (TPR) and False Positive Rate (FPR) are used to determine the class-specific performance of the models. TPR and FPR are calculated as follows:

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

where TP (true positive) indicates the number of cases for which a specific class occurred and the classifier predicted it occurred, TN (true negative) indicates the number of cases for which a specific class did not occur and the classifier predict that it did not occur, FP (false positive) indicates the number of cases for which a specific class did not occur but the classifier predicted that it occurred, and FN (false negative) indicates the number of cases for which a specific class occurred but the classifier predict that it did not occur. For example, if the classifier correctly predicted 132 of the 146 UC cases and predicted that 14 of them are LL, the TPR for UC is 132/(132 + 14) = 90.4%. Further, if the classifier predicted that 1 case from the 804 cases in other classes were UC, then the FPR for UC is 1/(1+803) = 0.1%.

3. RESULTS AND DISCUSSION

To investigate the performance of each machine learning method (KNN, RF, and LR) in classifying the normal mode and seven faulted modes of operation, all three methods were applied to the dataset. Figure 1 shows the estimated accuracy and true test accuracy for each method. Estimated accuracy is the 10-fold CV accuracy for the KNN and LR methods, while estimated accuracy is the OOB accuracy for the RF method. The figure shows that the RF and LR methods overestimated the true accuracy, while KNN method underestimated the true accuracy. Based on the true test accuracy, the LR method, with an overall accuracy rate of 93.6%, is the best classifier. The overall accuracy for RF and KNN methods are 88.3% and 83.6%, respectively.

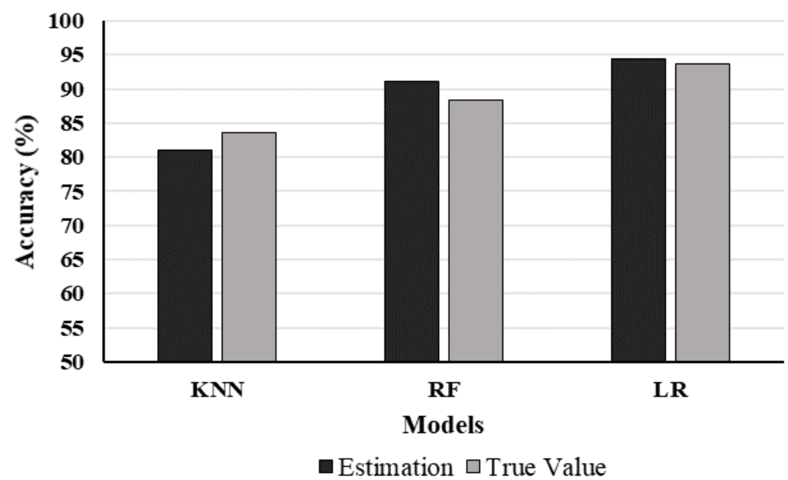

Figure 1: Estimated and true test accuracy for different classification methods

Confusion matrices for two algorithms, LR (highest accuracy) and KNN (lowest accuracy), are shown in Figure 2. The diagonal elements of the confusion matrices show the number of correct predictions, while off-diagonal elements show the incorrect predictions and characterize the nature of the error.

The confusion matrix on Figure 2(a) shows that 14 UC samples, 3 VL samples, 1 CA sample, 8 EA samples, 7 NC samples, and 17 NF samples are misclassified as members of class LL with KNN method. According to the confusion matrix in Figure 2(b), 5 UC samples, 6 OC samples, 3 VL samples, 2 EA samples, 3 NC samples, and 17 NF samples are misclassified as members of class LL with LR method. One very troubling aspect of these results is that all of the unfaulted tests (NF) are classified as faulted, for both classifiers. The reason for this is discussed below. KNN also struggles with the NC class, misdiagnosing NC cases as CA, LL, NF, and OC. Both classifiers frequently misclassify faults as LL faults, suggesting that there is some aspect of the training data – as opposed to the classification methods – that influences this result. The LR classifier in Figure 2(b) also misclassifies OC faults somewhat evenly as other faults, including EA, LL, NC and VL. There are no cases with either classifier where UC and OC are misclassified as one another.

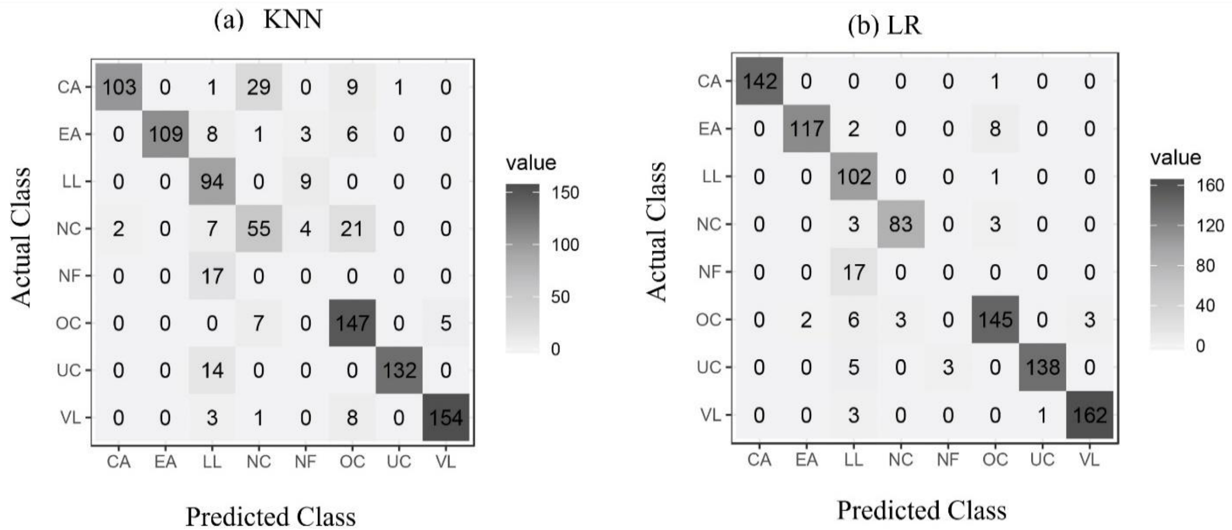

Figure 2: Confusion matrices for two classification methods: (a) KNN, (b) LR

Figures 3 and 4 illustrate the TPR and FPR values in each class of the faults for all three classification methods. A TPR of 100% means that all the samples of that class are correctly classified and an FPR of 0% means that none of the other classes are classified into that class. None of the classifiers could correctly predict the NF class samples (TPR of 0%). This poor performance is believed to be due to the imbalanced dataset used in this study. In our dataset only 48 samples out of the total 2851 samples belong to the NF class. The reason is that the matrix of test conditions includes combinations of operating conditions with fault intensities. While there are multiple fault intensities for faulted cases, the NF case has only one intensity – zero. Unfortunately, the NF case also may be hardest to classify, because all other classes resemble NF when they have low fault intensity. A final problem that compounds with these challenges for NF is that for practical deployment of FDD, the NF case is typically the most important to accurately classify, because any false alarm – classification of NF as a fault class – will result in a costly maintenance visit that is unnecessary and brings no performance benefit (Yuill and Braun, 2017).

To address the problem of imbalanced data, in ongoing work we are using an over-sampling approach in which new samples of the NF class are generated. For example, one method for doing this, the synthetic minority over-sampling technique (Chawla *et al*., 2002) generates new data using the K nearest neighbor of each of the NF cases. Another possible approach to address this problem is to augment the data set with additional data that contain small fault intensity faults – below some appropriate threshold – for each fault type, and to categorize these data as NF in the training set.

Figure 3 shows that the LR classification method has a very high TPR for all the classes except the NF class. The confusion matrix on Figure 2(b) shows that all 17 NF samples are misclassified as LL fault. This misclassification is also reflected in Figure 4, in which the LL fault has high FPR values. The UC fault has the lowest FPR value, followed by VL and EA faults. The LR classifier significantly outperforms the other classes in FPR rate for most faults, has a low FPR for all the classes except LL, and to a lesser extent OC.

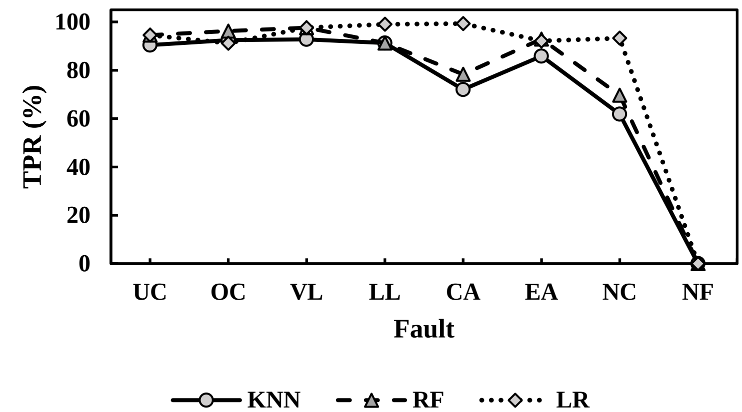

Figure 3: TPR values for each class for different classification methods

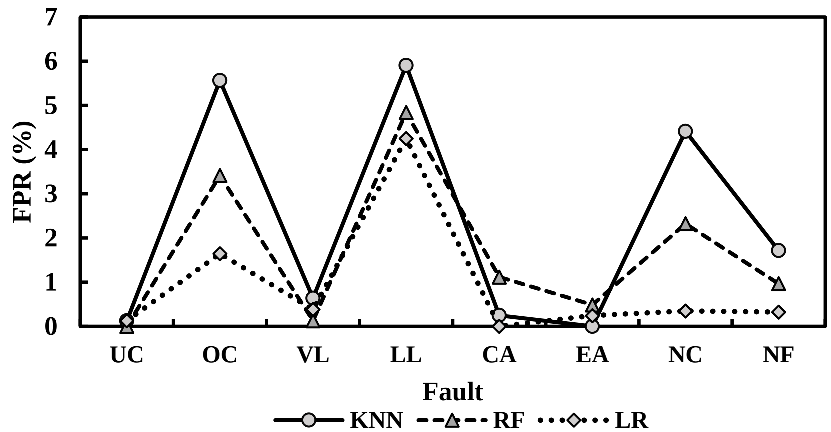

Figure 4: FPR values for each class for different classification methods

The ultimate goal for machine learning based fault classification is to provide a low cost and accurate diagnostic method that could be included onboard the RTU by the manufacturer. A significant amount of engineering research and development is needed before this is practical. For example, there needs to be a way to provide appropriate training data sets, the cost needs to be minimized by removing sensors that are not necessary, and cost-effective detection thresholds need to be determined. Thus, the work presented here is the first step, and clearly further investigation into the causes of different results for different faults and machine learning algorithms is needed. The characteristics of the training data play an important role in the potential for successful classification, so improvements in the fault distribution, including the magnitude of faults, within the training data is also needed. However, this work does succeed in showing that machine learning has potential to be used successfully to classify faults in RTU.

4. CONCLUSIONS

A data-driven RTU fault detection and diagnosis strategy is presented in this study. The proposed strategy formulates the FDD task as a multi-class classification problem. Three statistical machine learning classification methods are applied to a dataset in order to detect and diagnose the seven typical faults in RTU systems, using fifteen input variables. The results show that the classification algorithms can successfully detect and diagnose the faults in RTU systems. Based on our results, the following major conclusions can be drawn:

- The LR classification method, with an overall accuracy rate of 93.6%, is the best classifier of the three, and the KNN classification method, with an overall accuracy rate of 83.6%, is the worst classifier.

- None of the classification methods could correctly predict the NF class samples (TPR of 0%). This is because our original dataset is highly imbalanced, with only 48 samples out of the total 2851 samples belonging to the NF class (minority class). This is an important result of this work, because this problem is general – it will apply to future efforts to generate machine-learning based FDD from simulation data – and because the NF class is such an important class.

Overall, machine learning based FDD shows sufficient potential for further investigation. Future work to build upon these results should include:

- Generation of balanced data sets (by oversampling the NF class, for example)
- Application to data sets from additional RTU, to test how generalizable the resulting classifications are
- Study of the tradeoffs between the number of types of input (temperatures, pressures, etc.), and the effectiveness of the classifier
- Consideration of multiple simultaneous faults in the dataset as additional categories. Many FDD tools struggle with accurate diagnosis when multiple faults are present, so it would be beneficial to know whether machine learning based FDD has the potential to be more effective than status quo methods. Newly available data from tests with multiple simultaneous faults (Hu *et al.*, 2021; Hu and Yuill, 2021) may facilitate development in this area.
- Changes or tuning of fault intensity thresholds. Some of the fault levels in the training data set may not be severe enough to warrant the cost of repairing. These cases could be removed or reclassified as unfaulted for training purposes. This step could potentially help to address a shortcoming of the proposed classification-based FDD method, which is that it does not provide a fault severity assessment.

Although the LR classifier outperformed the other methods, it is difficult to predict whether this would continue to be the case when improvements to the FDD are made, by augmenting the training dataset to address NF diagnosis problems, for example. This work has shown that the classification method can have very significant impacts on the performance, so future work should continue to test multiple classification methods as improvements to the machine learning based FDD approach are made.

This investigation is an early exploration of the potential for machine learning FDD. The most significant challenge with data-driven methods is the high cost of developing training data, and it is possible that this will be an insurmountable barrier to widescale deployment of these methods. Nevertheless, until we fully understand the available potential, it is unknown in which circumstances they can provide practical benefit.

NOMENCLATURE

CA	condenser fouling
CV	cross-validation
EA	evaporator fouling
FDD	fault detection and diagnosis
FI	fault intensity
FN	false negative
FP	false positive
FPR	false positive rate
FXO	fixed orifice
k	counter for number of folds
K	counter for number of neighbors
KNN	K-nearest neighbors
LL	liquid-line restriction
LR	logistic regression
m	mass
$\dot{m}$	mass flow rate
N_2	nitrogen gas
NC	non-condensable gas
NF	no fault
OAR	overall accuracy rate
OC	refrigerant overcharge
OOB	out-of-bag
P	pressure
Power	power
RF	random forests
RTU	rooftop unit
T	dry bulb temperature
TN	true negative
TP	true positive
TPR	true positive rate
TXV	thermostatic expansion valve
UC	refrigerant undercharge
$\dot{V}$	volumetric flow rate
VL	compressor valve leakage
WB	wet bulb temperature

Subscripts

air	air
amb	ambient
c	condenser
ce	condenser exit
comp	compressor
dischg	discharge
e	evaporator
LL	liquid-line
RA	return air
SA	supply air
sat	saturation
suc	suction

Greek Symbols

Δ	difference

REFERENCES

Armstrong, P. R., Laughman, C. R., Leeb, S. B., & Norford, L. K. (2006). Detection of rooftop cooling unit faults based on electrical measurements. *HVAC&R Research*, *12*(1), 151–175. https://doi.org/10.1080/10789669.2006.10391172

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Braun, J. E. (1999). Automated fault detection and diagnostics for the HVAC&R industry. *HVAC&R Research*, *5*(2), 85–86. https://doi.org/10.1080/10789669.1999.10391225

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, B., & Braun, J. E. (2001). Simple Rule-Based Methods for Fault Detection and Diagnostics Applied to Packaged Air Conditioners. *ASHRAE Transactions*, 847–857.

Cheung, H., & Braun, J. E. (2013a). Simulation of fault impacts for vapor compression systems by inverse modeling. Part I: Component modeling and validation. *HVAC&R Research*, *19*(7), 892–906. https://doi.org/10.1080/10789669.2013.824800

Cheung, H., & Braun, J. E. (2013b). Simulation of fault impacts for vapor compression systems by inverse modeling. Part II: System modeling and validation. *HVAC&R Research*, *19*(7), 907–921. https://doi.org/10.1080/10789669.2013.819769

Comstock, M.C., Braun, J.E., & Groll, E.A. (2001). The sensitivity of chiller performance to common faults. *HVAC&R Research*, *7*(3), 263-279. DOI: 10.1080/10789669.2001.10391274

Ebrahimifakhar, A., Kabirikopaei, A., & Yuill, D. (2020). Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods. *Energy and Buildings*, *225*, 110318. https://doi.org/10.1016/j.enbuild.2020.110318

EIA (U.S. Energy Information Administration) (2012). Commercial Buildings Energy Consumption Survey.

https://www.eia.gov/consumption/commercial/.
Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181.
Han, H., Gu, B., Kang, J., & Li, Z. R. (2011). Study on a hybrid SVM model for chiller FDD applications. *Applied Thermal Engineering*, *31*(4), 582–592. https://doi.org/10.1016/j.applthermaleng.2010.10.021
Hu, Y., Yuill, D.P., Rooholghodos, A., Ebrahimifakhar, A., & Chen, Y. (2021). Impacts of simultaneous operating faults on cooling performance of a high efficiency residential heat pump. *Energy and Buildings. (In Press).*
Hu, Y., & Yuill, D.P. (2021). Effects of multiple simultaneous faults on characteristic fault detection features of a heat pump in cooling mode. *(In Review).*
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
Li, H., & Braun, J. E. (2007a). A methodology for diagnosing multiple simultaneous faults in vapor-compression air conditioners. *HVAC&R Research*, *13*(2), 369–395. https://doi.org/10.1080/10789669.2007.10390959
Li, H., & Braun, J. E. (2007b). An overall performance index for characterizing the economic impact of faults in direct expansion cooling equipment. *International Journal of Refrigeration*, *30*(2), 299–310. https://doi.org/10.1016/j.ijrefrig.2006.07.026
Li, H., & Braun, J. E. (2007c). Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners. *International Journal of Refrigeration*, *30*(3), 546–564. https://doi.org/10.1016/j.ijrefrig.2006.07.024
Li, H., & Braun, J. E. (2007d). Economic evaluation of benefits associated with automated fault detection and diagnosis in rooftop air conditioners. *ASHRAE Transactions*, *113 PART 2*, 200–210.
Li, H., & Braun, J. E. (2009a). Development, evaluation, and demonstration of a virtual refrigerant charge sensor. *HVAC&R Research*, *15*(1), 117–136. https://doi.org/10.1080/10789669.2009.10390828
Li, H., & Braun, J. E. (2009b). Virtual refrigerant pressure sensors for use in monitoring and fault diagnosis of Vapor-Compression equipment. *HVAC&R Research*, *15*(3), 597–616. https://doi.org/10.1080/10789669.2009.10390853
R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Rossi, T. M., & Braun, J. E. (1997). A statistical, Rule-Based fault detection and diagnostic method for vapor compression air conditioners. *HVAC&R Research*, *3*(1), 19–37. https://doi.org/10.1080/10789669.1997.10391359
Shoukas, G., Bianchi, M., & Deru, M. (2020). Analysis of fault data collected from automated fault detection and diagnostic products for packaged rooftop units. United States. doi:10.2172/1660228
Yan, R., Ma, Z., Zhao, Y., & Kokogiannakis, G. (2016). A decision tree based data-driven diagnostic strategy for air handling units. *Energy and Buildings*, *133*, 37–45. https://doi.org/10.1016/j.enbuild.2016.09.039
Yuill, D. P., & Braun, J. E. (2012). Evaluating fault detection and diagnostics protocols applied to air-cooled vapor compression air-conditioners. *Proceedings of 14th International Refrigeration and Air Conditioning Conference*, Paper 2470, West Lafayette, July 2012.
Yuill, D. P., & Braun, J. E. (2013). Evaluating the performance of FDD protocols applied to air-cooled unitary air-conditioning equipment. *HVAC&R Research*, *19*(7), 882-891. https://doi.org/10.1080/10789669.2013.808135
Yuill, D. P., & Braun, J. E. (2016). Effect of the distribution of faults and operating conditions on AFDD performance evaluations. *Applied Thermal Engineering, 106,* 1329–1336. https://doi.org/10.1016/j.applthermaleng.2016.06.149
Yuill, D. P., & Braun, J. E. (2017). A figure of merit for overall performance and value of AFDD tools. *International Journal of Refrigeration*, *74*, 649–659. https://doi.org/10.1016/j.ijrefrig.2016.11.015
Yuill, D. P., Cheung, H., & Braun, J. E. (2014). Validation of a fault-modeling equipped vapor compression system model using a fault detection and diagnostics evaluation tool. *Proceedings of 15th International Refrigeration and Air Conditioning Conference*, Paper 2606, West Lafayette, July 2014.
Yun, W. S., Hong, W. H., & Seo, H. (2021). A data-driven fault detection and diagnosis scheme for air handling units in building HVAC systems considering undefined states. *Journal of Building Engineering*, *35*, 102111. https://doi.org/10.1016/j.jobe.2020.102111
Zhao, Y., Wang, S., & Xiao, F. (2013). Pattern recognition-based chillers fault detection method using Support Vector Data Description (SVDD). *Applied Energy*, *112*, 1041–1048. https://doi.org/10.1016/j.apenergy.2012.12.043